

A Method of Record Linkage

by Akira Oshima,* Fumio Sakagami,* Aya Hanai,*
and Isaburo Fujimoto*

In cancer epidemiology, prospective approaches are very important both in testing etiological hypotheses and in evaluating preventive procedures. Prospective studies, however, are very difficult and expensive, because a large number of people and a long period of observation are necessary for a satisfactory study.

As a data source for follow-up studies, population-based cancer registry is very useful. The Osaka Cancer Registry has been in operation since December, 1962. Since 1968 the data processing, including the work of collation, has been semicomputerized. In order to identify cancer patients, we use the following six indices: date of birth, first Chinese character of a person's family name, address a: city, ward, town or village, address b: further details, i.e., street, avenue, section, hamlet etc., site, and sex.

When we have data on the collation indices for the subjects to be followed up, we can conduct follow-up studies easily and accurately, using a semicomputerized collation method similar to that in the cancer registration system. Because the master file of the Osaka Cancer Registry contains the data of cancer cases reported and all cancer deaths among the residents of Osaka Prefecture, we can follow up the subjects living in Osaka Prefecture and obtain data about virtually all cancer incidences and deaths among them. In this follow-up method by means of record linkage to the cancer registry, some considerations should be taken into account for the following factors; coverage of cancer data in the Osaka Cancer Registry, reliability of the collation method, and address of the subjects to be followed up.

As an example of a study with this method, we present the follow-up study of the screenees of a mass screening program for stomach cancer.

Introduction

In cancer epidemiology, prospective approaches are very important both in testing etiological hypotheses and in evaluating preventive procedures (1). Prospective studies, however, are very difficult and expensive, because a large number of people and a long period of observation are necessary for a satisfactory study. For a follow-up study in Japan we have three kinds of data sources, the systems of which are routinely maintained for other purposes.

First, there is a system of official family records, the Koseki, containing the Honseki, i.e., the legal address. By knowing the recent Koseki record of a person, we can obtain information on his survival status: whether he is alive or dead. But it takes a great deal of work to carry out an inquiry into Koseki records one by one for a large number of people, because the Koseki system has not yet been computerized.

Second, the health centers in Japan routinely collect and keep a copy of the death certificate for every dead person whose address comes within their jurisdiction. These data are useful for a follow-up study. Since 1969 the data from death certificates have been transferred on to magnetic tapes with the optical mark reader in the Health and Welfare Statistics, Information Department of the Ministry of Health and Welfare. Unfortunately, however, the name of a dead person is not included in the items for input into the tape, and we have no identification number in Japan, such as the National Health Service number in the United Kingdom and the Social Security number in the United States. As a result, follow-up studies by collating the file of subjects to be followed up and the mortality data file with a computer are impossible. It takes a great deal of work to follow up a large number of people one by one, using the copy of the death certificate which is kept in the health center.

As a third data source for a follow-up study we have a cancer registry system. As of 1977, eleven prefectural cancer registries are in active operation.

*Osaka Cancer Registry, Department of Field Research, the Center for Adult Diseases, 3 Nakamichi 1-chome, Higashinari-ku, Osaka, Japan.

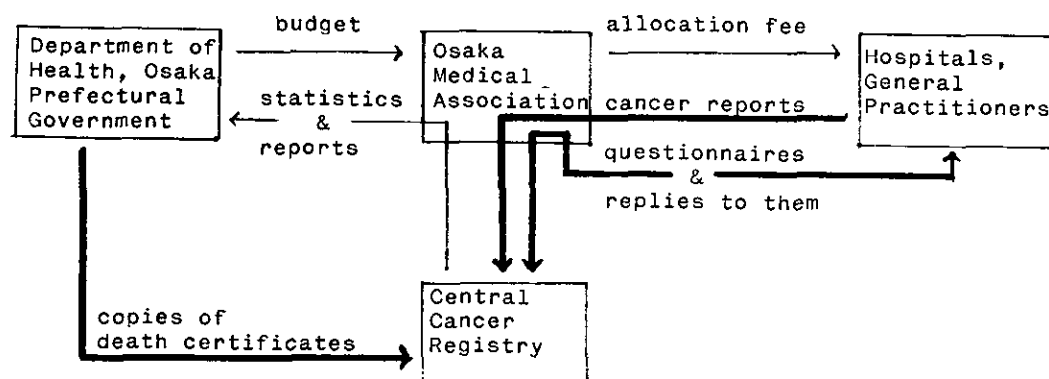


FIGURE 1. System of cancer registration in Osaka.

Though most of the registries have very limited capacities because of their insufficient budget and staffing, some registries are now capable of conducting various follow-up studies using their registered data.

In this paper, we introduce, first, the cancer registration system in Osaka, and, then, the follow-up method by means of semicomputerized record linkage to the cancer registry. This follow-up system has recently been set up by the Osaka Cancer Registry (2).

Cancer Registration System in Osaka

Outline of the Osaka Cancer Registry

Cancer registration in Osaka (population: 8.3 million as of 1975) has been in operation since December, 1962. The objectives of registration are as follows: (1) to provide information concerning the size and nature of the cancer problem in Osaka and to present data for cancer epidemiology; (2) to provide information concerning the medical care and prognosis of cancer patients and to assist in making cancer control projects and in evaluating their effectiveness; and (3) to return its statistics and informations about each patient's prognosis to each participating hospital and clinic and to assist in raising the level of their medical care for cancer patients.

Figure 1 shows the system of the Osaka Cancer Registry. Hospital doctors and general practitioners send cancer reports to the Osaka Medical Association. The O.M.A. transfers these cancer reports monthly to the central registry, which has been set up in the Department of Field Research of the Center for Adult Diseases, Osaka. The Department of Health of the Osaka Prefectural Government sends Xerox copies of all death certificates to the central registry. At the central registry, those copies that indicate the presence of cancer are transferred on to

cancer death cards. Some cancer death cards remain unmatched in the collation between cancer reports and cancer death cards. To the doctors who signed these death certificates, the registry sends questionnaires on the patients' medical care.

All the cancer reports and cancer death cards are filed in order of the patient's birth date. When there are two or more cards for one patient, all these cards are stapled together in order of the date of diagnosis. Table 1 shows the number of cancer reports and cancer death cards which have been collected at the central registry.

Semicomputerized Cancer Registration System

It is an important job of the central registry to collate all cancer reports and cancer death cards in order to eliminate duplicate registration. Before 1967 the collation work was done manually; first, by

Table 1. Number of cancer reports and cancer death cards.

| Year | Number of cards | | |
|------|--------------------------------------|-------------------------------------|--|
| | Cancer reports by year of acceptance | Cancer death cards by year of death | Replies to questionnaires by year of death |
| 1963 | 6,955 | 6,235 | 1,944 |
| 1964 | 5,696 | 6,612 | 2,369 |
| 1965 | 8,256 | 6,749 | 2,620 |
| 1966 | 9,364 | 6,962 | 2,117 |
| 1967 | 9,491 | 7,207 | 1,366 |
| 1968 | 11,481 | 7,523 | 2,295 |
| 1969 | 10,410 | 7,629 | 1,934 |
| 1970 | 12,081 | 8,211 | 1,580 |
| 1971 | 10,471 | 8,669 | 2,585 |
| 1972 | 11,626 | 8,985 | 2,437 |
| 1973 | 12,039 | 9,381 | 2,604 |
| 1974 | 12,727 | 9,501 | 2,435 |
| 1975 | 11,875 | 9,987 | 2,433 |
| 1976 | 15,185 | 10,484 | — |
| 1977 | 15,179 | — | — |

sorting the two files to be collated in order of the patient's date of birth, and then by sorting them again in order of the patient's address. But because of the increasing number of cases registered, the collation work has become too large for a small registry staff. In 1968 the registry decided to computerize all tasks, including the collation work.

Figure 2 shows the computer system flow chart for collation, now in operation. In Step 1, the new report file, collations are made against each other to check duplication; in Step 2, the new report file in which duplication has been checked at step 1 is collated with the master file, and these two files are then combined into a new master file; in Step 3, a new cancer death file is collated with the master file in order to check the survival status of the previously registered cases and to check the cases registered by death certificate only.

In Osaka, there are five medical schools, 456 hospitals, and 6077 private clinics; 11,423 physicians were working as of 1975. With so many participants in the cancer registrations, there is a risk of misinformation being included on indices submitted for collation. Thus, we decided to use combinations of several indices for identification of cancer cases. The following six indices are used; (1) date of birth, (2) first Chinese character of a person's family name, (3) address a (city, ward, town or village), (4) address b (further details, i.e., street, avenue, section, hamlet), (5) collation site, and (6) sex.

In Osaka, a post card has been used as a cancer report form. Sometimes doctors write only a part of the patient's name out of considerations of confidence. Therefore, only the first Chinese character of the patient's family name can be used for identifica-

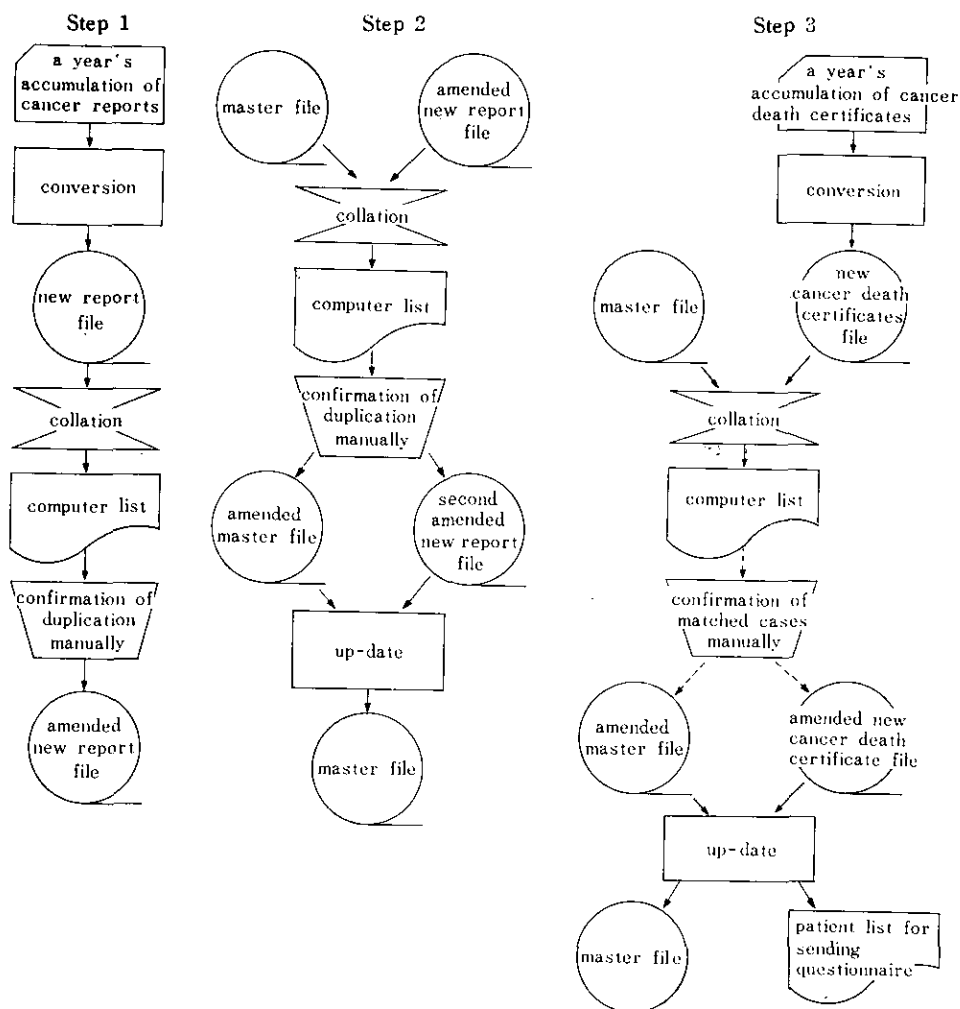


FIGURE 2. Computer system flow chart in the central cancer registry.

Table 2. Combinations of the six indices' agreement or disagreement, selected as possible matching pairs.

| Tape copy number | Combination number | Collation index ^a | | | | | |
|------------------------|-----------------------|------------------------------|------|--------------|--------------|-------------------|-----|
| | | Birth date | Name | Address a | Address b | Collation site | Sex |
| 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 2 | 0 | 0 | 0 | X | 0 | 0 |
| | 3 | 0 | 0 | 0 | 0 | X | 0 |
| | 4 | 0 | 0 | 0 | X | X | 0 |
| | 5 | 0 | 0 | X | X | 0 | 0 |
| | 6 | 0 | X | 0 | 0 | 0 | 0 |
| | 7 | 0 | X | 0 | X | 0 | 0 |
| | 8 | 0 | 0 | 0 | 0 | 0 | X |
| | 9 | 0 | 0 | 0 | X | 0 | X |
| | 12 | 0 | 0 | 0 | 0 | X | X |
| | 13 | 0 | 0 | 0 | X | X | X |
| | 14 | 0 | 0 | X | X | 0 | X |
| | 15 | 0 | 0 | X | X | X | 0 |
| | 16 | 0 | X | 0 | 0 | 0 | X |
| | 17 | 0 | X | 0 | X | 0 | X |
| | 18 | 0 | X | 0 | 0 | X | 0 |
| | 19 | 0 | X | 0 | X | X | 0 |
| | 26 | 0 | X | X | X | 0 | 0 |
| | 28 | 0 | 0 | X | X | X | X |
| | 29 | 0 | X | 0 | 0 | X | X |
| | 30 | 0 | X | 0 | X | X | X |
| 2 | 10 | X | 0 | 0 | 0 | 0 | 0 |
| | 11 | X | 0 | 0 | X | 0 | 0 |
| | 20 | X | 0 | 0 | 0 | 0 | X |
| | 21 | X | 0 | 0 | X | 0 | X |
| | 22 | X | 0 | 0 | 0 | X | 0 |
| 3 | 23 | X | 0 | 0 | X | X | 0 |
| | 24 | X | X | 0 | 0 | 0 | 0 |
| | 25 | X | X | 0 | X | 0 | 0 |

^a(0) agreement; (X) disagreement.

tion index. Each Chinese character is given its own particular pronunciation and encoded under letters of the Roman alphabet.

A collation site code is added automatically by a computer to every cases according to the primary site code. For example, to all cases whose primary site codes are 180-182 or 2340 (I.C.D.), 180 is given as the collation site code.

Table 2 shows the 29 combinations of the agreement or disagreement of the six indices, which were selected empirically as those of possible matching pairs.

All pairs coming under these combinations are printed out on the computer list. The detailed flow chart of collation is shown in Figure 3. All the possible pairs printed out by the computer are examined by a manual check of the original cards in order to determine whether or not the two data belong to the same patient.

Table 3 shows the number of pairs printed out and the number of cases which were confirmed manually by their combination number when we checked the matching efficiency using the data for 1968. From

this result, we decided to discard thereafter combinations 21, 23, 24, and 25, and to add conditions for some combination numbers. The reason for this is that the date of birth is divided into four parts: (1) the era name of the emperor when the patient was born, (2) year of birth according to the Japanese calendar, (3) month, and (4) date. For combination numbers 11, 20, and 22, the possible matching pairs are to be printed out only when three of the four factors of their birth dates agree.

Pairs of combination number 11, 20, and 22 are to be printed out when the birth date or address b of one card is not given. For combination number 26, pairs are to be printed out only when the name of one card is not given.

As a result of these modifications, the number of pairs printed out on the list in step 3 with the 1968 data decreased to 39.5% of those on the previous computer list, and the number of confirmed cases remained 98.9% of those before the modifications (Table 4).

In order to check the reliability, both the manual method of collation and the computer method were

used in the collation of the cancer reports diagnosed in 1968 with the cancer death cards of 1968. As shown in Table 5, there were 2710 confirmed matching pairs by either method. The computer method overlooked 2.4% of all the matching pairs, which was a little higher than the percentage of pairs that was overlooked by the manual method (1.2%).

But because of the increased number of data to be collated, the manual method is now impracticable for a small registry staff. Instead, by the semicomputerized method described above, it has become possible to maintain a high degree of reliability for all collations even when a large number of data is to be collated.

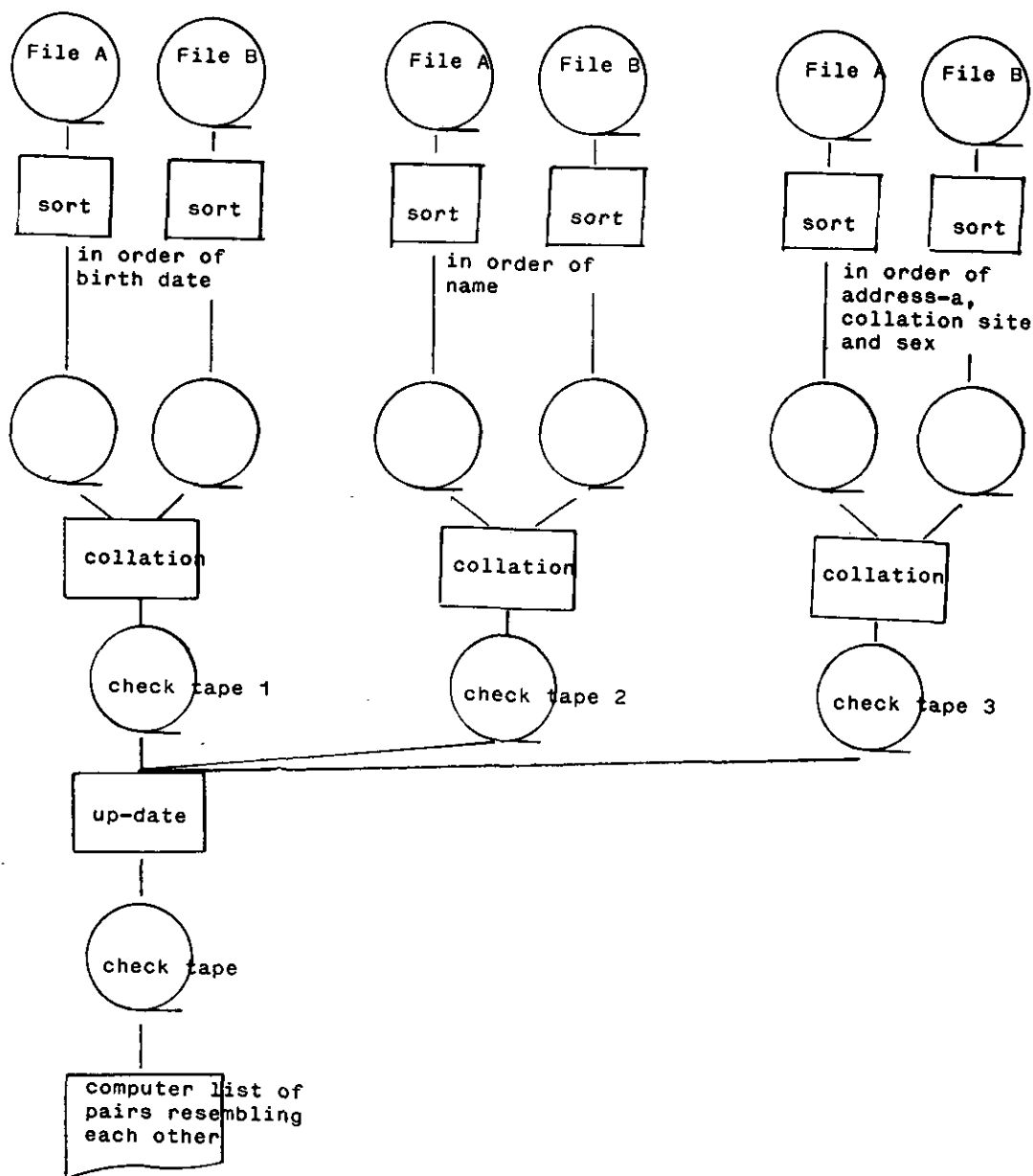


FIGURE 3. Flow chart of collation: (step 1) file A: year's cancer report file; file B: year's cancer report file same as file A); (step 2) file A: master file; file B: year's cancer report file; (step 3) file A: master file; file B: year's cancer death file.

Table 3. Confirmation rate by combination number.

| Combination number | Step 1 | | | Step 2 | | | Step 3 | | |
|--------------------|------------------------------|----------------------------|--------------|------------------------------|----------------------------|--------------|------------------------------|----------------------------|--------------|
| | No. of pairs on the list (A) | No. of confirmed cases (B) | B/A X 100, % | No. of pairs on the list (A) | No. of confirmed cases (B) | B/A X 100, % | No. of pairs on the list (A) | No. of confirmed cases (B) | B/A X 100, % |
| 1 | 1,358 | 1,358 | 100.0 | 416 | 416 | 100.0 | 1,645 | 1,645 | 100.0 |
| 2 | 121 | 121 | 100.0 | 44 | 44 | 100.0 | 215 | 215 | 100.0 |
| 3 | 144 | 144 | 100.0 | 42 | 42 | 100.0 | 179 | 179 | 100.0 |
| 4 | 15 | 15 | 100.0 | 5 | 5 | 100.0 | 31 | 30 | 96.8 |
| 5 | 53 | 44 | 83.1 | 29 | 22 | 75.9 | 110 | 96 | 87.3 |
| 6 | 119 | 116 | 97.5 | 36 | 36 | 100.0 | 105 | 102 | 97.1 |
| 7 | 41 | 22 | 53.7 | 14 | 6 | 42.9 | 46 | 15 | 32.6 |
| 8 | 8 | 8 | 100.0 | 2 | 2 | 100.0 | 23 | 23 | 100.0 |
| 9 | 0 | — | — | 1 | 1 | 100.0 | 5 | 5 | 100.0 |
| 12 | 0 | — | — | 1 | 1 | 100.0 | 2 | 2 | 100.0 |
| 13 | 0 | — | — | 1 | 0 | 0 | 1 | 0 | 0 |
| 14 | 3 | 0 | 0 | 3 | 1 | 33.3 | 15 | 4 | 26.7 |
| 15 | 37 | 5 | 13.5 | 11 | 2 | 18.2 | 38 | 11 | 28.9 |
| 16 | 1 | 0 | 0 | 0 | — | — | 5 | 5 | 100.0 |
| 17 | 4 | 0 | 0 | 1 | 0 | 0 | 18 | 0 | 0 |
| 18 | 17 | 15 | 88.2 | 6 | 4 | 66.7 | 17 | 13 | 76.5 |
| 19 | 33 | 3 | 9.1 | 17 | 0 | 0 | 66 | 4 | 6.1 |
| 26 | 621 | 8 | 1.3 | 308 | 1 | 0.3 | 929 | 2 | 0.2 |
| 28 | 24 | 0 | 0 | 9 | 0 | 0 | 34 | 0 | 0 |
| 29 | 1 | 0 | 0 | 0 | — | — | 3 | 0 | 0 |
| 30 | 47 | 0 | 0 | 36 | 0 | 0 | 73 | 0 | 0 |
| 10 | 422 | 331 | 78.4 | 78 | 50 | 64.1 | 984 | 721 | 73.3 |
| 11 | 2,967 | 31 | 1.0 | 810 | 12 | 1.5 | 4,715 | 51 | 1.1 |
| 20 | 57 | 1 | 1.8 | 24 | 0 | 0 | 146 | 3 | 2.1 |
| 22 | 304 | 39 | 12.8 | 90 | 4 | 5.6 | 561 | 60 | 10.7 |
| Subtotal | 6,397 | 2,261 | 35.3 | 1,984 | 649 | 32.7 | 9,966 | 3,186 | 32.0 |
| 21 | 2,900 | 0 | 0 | | | | — | 0 | — |
| 23 | 18,575 | 1 | 0.0 | | | | — | 8 | — |
| 24 | 15,450 | 13 | 0.0 | | | | — | 36 | — |
| 25 | 5 | 0 | 0 | | | | — | 0 | — |
| Grand total | 43,327 | 2,275 | 5.3 | — | — | — | — | — | — |
| none ^a | — | 9 | — | | | | — | 33 | — |

^aThe pairs for which no combination numbers were given were not printed out on the list.

Table 4. Results of program modification^a

| Combination number | Modification for | No. of pairs on the list (A) | | No. of confirmed cases (B) | | B/A X 100, % | |
|--------------------|------------------|------------------------------|--------------|----------------------------|--------------|---------------|--------------|
| | | Before modif. | After modif. | Before modif. | After modif. | Before modif. | After modif. |
| 11 | Birth date | 4,715 | 184 | 51 | 39 | 1.2 | 21.2 |
| 20 | Birth date | 146 | 4 | 3 | 2 | 2.7 | 50.0 |
| 22 | Birth date | 561 | 48 | 60 | 39 | 10.7 | 81.3 |
| | Address-b | | | | | | |
| 26 | Name | 929 | 40 | 2 | 2 | 0.2 | 5.0 |
| All others | None | 3,615 | 3,615 | 3,070 | 3,070 | 84.9 | 84.9 |
| Total (%) | | 9,966 (100.0) | 3,891 (39.5) | 3,186 (100.0) | 3,152 (98.9) | 32.0 | 81.0 |

^aThese data are the results of collation in step 3 with the 1968's data.

Application

Follow-Up Method by Means of Record Linkage to Cancer Registry

Using the semicomputerized collation method similar to that in the cancer registration system, we can conduct various follow-up studies when we have data on the collation indices for the subjects to be followed up. Figure 4 shows the follow-up method by means of record linkage to the Osaka Cancer Registry. Because the master file of the registry contains the data of cancer cases reported and all cancer deaths among the residents of Osaka Prefecture, we can follow up all the subjects living in Osaka Prefecture and obtain data about virtually all cancer incidences and deaths occurring among them.

By this method we can conduct follow-up studies for a large number of subjects easily and accurately for a long period, but the following factors should be kept in mind.

Coverage of Cancer Data in the Osaka Cancer Registry. By this method we cannot obtain data about cancer incidences or deaths of persons where there were no reports to the registry even if they suffered from cancer or died of it. The coverage of cancer incidence data in the registry is dependent on the degree of participation in the registration by each hospital or physician. In 1966, we estimated that the coverage of cancer incident cases in the registry was about 90%. As for cancer deaths, we can collect all data from the Department of Health of the Osaka Prefectural Government.

Reliability of the Collation Method. As described above, the percentage of pairs that are overlooked by this semicomputerized method is about 2%.

Address of the Subjects to Be Followed Up. The Osaka Cancer Registry contains the data of cancer incidences or deaths occurring only among the residents of Osaka Prefecture. Therefore, we can follow up by this method only those subjects whose addresses fall within the prefecture. When the subjects move outside the prefecture during the follow-up period, we cannot obtain cancer data about them and they are regarded as being free of cancer.

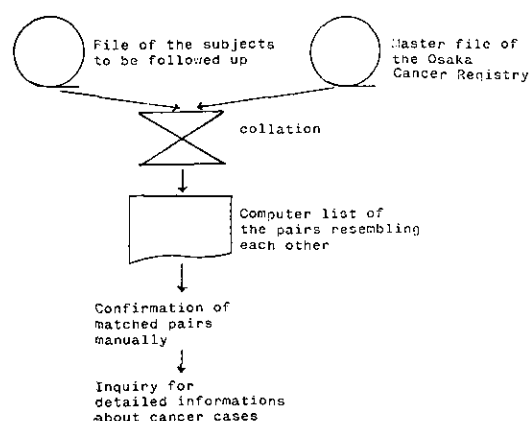


FIGURE 4. The follow-up method by means of record linkage to the Osaka Cancer Registry.

In analyzing data obtained from this follow-up study, special considerations should be given to these factors.

Follow-Up Study of the Screenees of a Mass Screening Program for Stomach Cancer

As an example of a follow-up study by means of record linkage to the cancer registry, we describe in this section a follow-up study of the screenees of a mass screening program for stomach cancer.

In evaluating the screening program, it is necessary to follow up all the screenees and to show the decrease in mortality from stomach cancer among them (3). But the actual follow-up of a large number of screenees requires a great deal of work. Instead, we followed up the screenees by means of record linkage to the Osaka Cancer Registry.

The study subjects are the 32,789 screenees of the stomach cancer screening tests conducted by the Center for Adult Diseases, Osaka between 1967 and 1970. Those who lived outside Osaka Prefecture and whose stomach had been resected before the initial screening are excluded from the study subjects. The screening was carried out with a specially designed mobile unit equipped with a photofluorographic apparatus. As shown in Table 6, about 20% of the total

Table 5. Comparison of the computer method with the manual method.^a

| | No. | % |
|--|------|-------|
| Confirmed by either the computer method or the manual method | 2710 | 100.0 |
| Overlooked by the computer method | 64 | 2.4 |
| Overlooked by the manual method | 32 | 1.2 |

^aThese data are the results of collation of the cancer reports diagnosed in 1968 with the cancer death cards of 1968. Those overlooked by the computer method are due to: (a) combination numbers not given, (b) disuse of certain combination numbers and (c) addition of certain conditions to certain combination numbers.

Table 6. Results of the initial screening among the study subjects.

| | Male | | Female | | Total | |
|----------------------------------|--------|---------|--------|---------|--------|---------|
| | No. | (%) | No. | (%) | No. | (%) |
| Total screened | 18,578 | (100.0) | 14,211 | (100.0) | 32,789 | (100.0) |
| Further examinations recommended | 5,317 | (28.6) | 2,329 | (16.4) | 7,646 | (23.3) |
| Further examinations performed | 4,459 | (24.0) | 1,930 | (13.6) | 6,389 | (19.5) |
| Stomach cancer cases detected | | | | | | |
| "Early" cancer cases | 19 | (0.10) | 12 | (0.08) | 31 | (0.09) |
| Advanced cancer cases | 60 | (0.32) | 32 | (0.23) | 92 | (0.28) |
| Total | 79 | (0.42) | 44 | (0.31) | 123 | (0.37) |

screenees were recommended for further examinations, including roentgenography, endoscopy, cytology and/or biopsy. A total of 123 stomach cancer cases were thus detected. Of these, 31 were "early" stomach cancer cases. By linking the file of the screenees with that of the Osaka Cancer Registry, we obtained data about stomach cancer cases and deaths after the initial screening. The follow-up period was from the initial screening through December 31, 1975, and the average follow-up period was 6.1 years.

Table 7 shows the number of stomach cancer cases diagnosed during the follow-up period. In addition to the 123 stomach cancer cases detected at the initial screening, 212 further cases in total were diagnosed after the initial screening. Seventy cases were detected by the repeat screening and the remaining 142 cases were diagnosed outside the screening. The data of most cases diagnosed outside the screening were obtained for the first time by the record linkage to the cancer registry. The percentage of those who were screened only once during the follow-up period among the total screenees was 45.2% in males and 62.2% in females. The average number of screenings per screenee during the follow-up period was 2.66 in males and 1.92 in females.

The number of stomach cancer deaths among the study subjects during the follow-up period is shown in Table 8. There were 63 deaths among the 123 stomach cancer cases detected at the initial screen-

ing, 25 deaths among the 70 cases detected at the repeat screening and 91 deaths among the 142 cases not diagnosed by screening. In total, 179 stomach cancer deaths were observed among the study subjects during the follow-up period.

In order to obtain the expected number of deaths for comparison with the observed number of deaths, the person-years observed during the follow-up period were calculated by sex and age-group. In this calculation, consideration was given to such factors as aging and dropping out from observation due to death and moving out of Osaka Prefecture. The probability of death was obtained from the life table calculated on the basis of the 1970 mortality statistics. The rate of moving out of Osaka Prefecture was estimated from the 2% sample actually followed up, where the mail survey and Koseki survey were conducted. Then, the age-specific mortality rate of stomach cancer in the general population of Osaka Prefecture was applied to the person-years of each age-group. The stomach cancer mortality rates used were the average of those for 1969-1971 and those for 1972-1974, obtained from the Osaka Cancer Registry. We also calculated the expected numbers of stomach cancer cases, deaths from cancer of digestive organs and deaths from cancer of all sites by similar methods and compared them with the observed numbers in order to examine whether there are any biases in the study subjects.

Table 9 shows the comparison of the observed

Table 7. Stomach cancer cases diagnosed during the follow-up period.

| | Male ^a | Female ^a | Total ^a |
|--------------------------------|-------------------|---------------------|--------------------|
| Detected by repeat screening | 53 (20) | 17 (6) | 70 (26) |
| Diagnosed not due to screening | 98 (16) | 44 (6) | 142 (22) |
| Total | 151 (36) | 61 (12) | 212 (48) |

^aThe number in parenthesis expresses "early" stomach cancer cases.

Table 8. Stomach cancer deaths during the follow-up period^a

| | Male | Female | Total |
|--|---------|--------|---------|
| Deaths among cases detected at the initial screening | 49/ 79 | 14/ 44 | 63/123 |
| Deaths among cases detected by repeat screening | 18/ 53 | 7/ 17 | 25/ 70 |
| Deaths among cases diagnosed not due to screening | 62/ 98 | 29/ 44 | 91/142 |
| Total | 129/230 | 50/105 | 179/335 |

^aThe numerator expresses stomach cancer deaths and the denominator expresses stomach cancer cases.

number with the expected one when the study subjects are divided by age-groups at the initial screening. As for those aged 40 to 59 years, the O/E ratio in stomach cancer death was 0.72 in males, 0.79 in females and 0.74 in total. Namely, there was about a 20 to 30% decrease in mortality from stomach cancer in this age-group. This definitely demonstrates the effect of the screening program. The O/E ratio of stomach cancer cases was 1.25 in males, 1.57 in females and 1.35 in total. When only advanced stomach cancer cases were considered, the ratio was 0.94 in males, 1.20 in females and 1.02 in total. These figures show that there were about the same number of stomach cancer cases in the study subjects as in the general population. But the O/E ratio in death from cancers of digestive organs and in deaths from

cancers of all sites shows that there might be some biases in the study subjects aged 40 to 59.

Table 9 also shows the O/E ratios for stomach cancer deaths and cases among the study subjects of other age-groups. Though the observed numbers are small, the tendency can be seen for the effect of the screening program in reducing stomach cancer deaths to be more pronounced in younger age-groups than in older ones.

We are planning to continue this follow-up study for a fairly long period in the hope that the knowledge from it will be useful both in elucidating the natural history of stomach cancer and in answering practical questions such as the optimal frequency and interval for the screening program.

Table 9. Comparison of the observed number with the expected one, by age group.

| Age group | Item for comparison | Male | | | Female | | | Total | | |
|-----------|---|------|-------|------|--------|-------|------|-------|--------|------|
| | | O | E | O/E | O | E | O/E | O | E | O/E |
| ≤39 | Stomach cancer deaths | 2 | 5.35 | 0.37 | 4 | 4.15 | 0.96 | 6 | 9.50 | 0.63 |
| | Stomach cancer cases | 12 | 8.37 | 1.43 | 8 | 5.32 | 1.50 | 20 | 13.69 | 1.46 |
| | Stomach cancer cases (advanced cases) | 4 | 8.37 | 0.48 | 7 | 5.32 | 1.32 | 11 | 13.69 | 0.80 |
| | Deaths from cancer of the digestive organ, excluding stomach cancer | 1 | 2.54 | 0.39 | 6 | 1.46 | 4.11 | 7 | 4.00 | 1.75 |
| | Deaths from cancer of all sites, excluding stomach cancer | 4 | 8.83 | 0.45 | 10 | 8.66 | 1.15 | 14 | 17.49 | 0.80 |
| 40-59 | Stomach cancer deaths | 52 | 72.29 | 0.72 | 26 | 32.81 | 0.79 | 78 | 105.10 | 0.74 |
| | Stomach cancer cases | 118 | 94.53 | 1.25 | 64 | 40.72 | 1.57 | 182 | 135.25 | 1.35 |
| | Stomach cancer cases (advanced cases) | 89 | 94.53 | 0.94 | 49 | 40.72 | 1.20 | 138 | 135.25 | 1.02 |
| | Deaths from cancer of the digestive organ, excluding stomach cancer | 24 | 30.66 | 0.78 | 14 | 14.88 | 0.94 | 38 | 45.54 | 0.83 |
| | Deaths from cancer of all sites, excluding stomach cancer | 64 | 97.78 | 0.65 | 35 | 70.41 | 0.50 | 99 | 168.19 | 0.59 |
| 60-69 | Stomach cancer deaths | 55 | 43.78 | 1.26 | 16 | 19.02 | 0.84 | 71 | 62.80 | 1.13 |
| | Stomach cancer cases | 78 | 50.56 | 1.54 | 27 | 22.30 | 1.21 | 105 | 72.86 | 1.44 |
| | Stomach cancer cases (advanced cases) | 61 | 50.56 | 1.21 | 20 | 22.30 | 0.90 | 81 | 72.86 | 1.11 |
| | Deaths from cancer of the digestive organ, excluding stomach cancer | 16 | 17.54 | 0.91 | 10 | 9.63 | 1.04 | 26 | 27.17 | 0.96 |
| | Deaths from cancer of all sites, excluding stomach cancer | 39 | 57.11 | 0.68 | 24 | 33.08 | 0.73 | 63 | 90.17 | 0.70 |
| ≥ 70 | Stomach cancer deaths | 20 | 14.76 | 1.36 | 4 | 5.49 | 0.73 | 24 | 20.25 | 1.19 |
| | Stomach cancer cases | 22 | 16.07 | 1.37 | 6 | 5.97 | 1.01 | 28 | 22.04 | 1.27 |
| | Stomach cancer cases (advanced cases) | 21 | 16.07 | 1.31 | 5 | 5.97 | 0.84 | 26 | 22.04 | 1.18 |
| | Deaths from cancer of the digestive organ, excluding stomach cancer | 9 | 5.85 | 1.54 | 1 | 2.74 | 0.36 | 10 | 8.59 | 1.16 |
| | Deaths from cancer of all sites, excluding stomach cancer | 13 | 18.34 | 0.71 | 4 | 8.60 | 0.47 | 17 | 26.94 | 0.63 |

Summary

In this paper, we have introduced, first, the cancer registration system in Osaka, and then, the follow-up method by means of record linkage to the cancer registry.

As an example of a study with this method, we have next presented the follow-up study of the screenees of a mass screening program for stomach cancer.

By this method, we are now able to conduct various prospective studies, concurrent or nonconcurrent, in order to set up or to test etiological

hypotheses of cancer. We are planning to collect various lists of the population at risk, where the identification indices for collation are included.

REFERENCES

1. Lilienfeld, A. M. *Foundations of Epidemiology*. Oxford Univ. Press, New York, 1976.
2. Hanai, A., Sakagami, F., and Fujimoto, I. Computerized cancer collation system. *Ann. Repts. Center Adult Dis.* 13: 1 (1973).
3. Sackett, D. Periodic examination of patients at risk. In: *Cancer Epidemiology and Prevention*, D. Schottenfeld, Ed., Charles C Thomas, Springfield, Ill., 1975, pp. 437-454.